ELECTRONIC FRONTIER FOUNDATION
Protecting Rights and Promoting Freedom on the Electronic Frontier

The Honorable John Thune
U.S. Senate Committee on Commerce, Science, and Transportation, Chairman
512 Dirksen Senate Building
Washington, DC 20510

The Honorable Bill Nelson
U.S. Senate Committee on Commerce, Science, and Transportation, Ranking Member
716 Senate Hart Office Building
Washington, DC 20510

Chairman Thune, Ranking Member Nelson, and members of the U.S. Senate Committee on Commerce, Science, and Transportation:

The Electronic Frontier Foundation is the leading nonprofit organization defending civil liberties on the Internet. Founded in 1990, EFF champions user privacy, free expression, and innovation through impact litigation, policy analysis, grassroots activism, and technology development.

EFF strongly opposes the Stop Enabling Sex Traffickers Act (S. 1693). SESTA would be disastrous for freedom of speech and expression online as well as for innovation among online businesses. The bill would do very little or nothing to achieve its purpose of fighting sex trafficking, but it would significantly threaten legitimate online speech. We are particularly concerned that it would silence marginalized voices online, including those of sex trafficking victims.

We've included our four-part analysis of SESTA in which we identify the key ways in which its implementation would compromise civil liberties. We originally published these essays on our website. As you review SESTA, please use our analysis to consider how the bill would undermine your constituents' rights. We would be pleased to discuss these issues with you further.

Sincerely,

The Electronic Frontier Foundation

Enclosed:

- *Section 230 Is Not Broken*
- *Congress Doesn't Understand How Section 230 Works*
- *Stop SESTA: Amendments to Federal Criminal Sex Trafficking Law Sweep Too Broadly*
- *Whose Voices Will SESTA Silence?*

815 Eddy Street • San Francisco, CA 94109 USA
voice +1 415 436 9333 fax +1 415 436 9993 web www.eff.org email information@eff.org

Section 230 Is Not Broken

Sophia Cope, Staff Attorney

https://www.eff.org/deeplinks/2017/09/stop-sesta-section-230-not-broken

EFF opposes the Senate's Stop Enabling Sex Traffickers Act (S. 1693) ("SESTA"), and its House counterpart the Allow States and Victims to Fight Online Sex Trafficking Act (H.R. 1865), because they would open up liability for Internet intermediaries—the ISPs, web hosting companies, websites, and social media platforms that enable users to share and access content online—by amending Section 230's immunity for user-generated content (47 U.S.C. § 230). While both bills have the laudable goal of curbing sex trafficking, including of minor children, they would greatly weaken Section 230's protections for online free speech and innovation.

Proponents of SESTA and its House counterpart view Section 230 as a broken law that prevents victims of sex trafficking from seeking justice. But Section 230 is not broken. First, existing federal criminal law allows federal prosecutors to go after bad online platforms, like Backpage.com, that knowingly play a role in sex trafficking. Second, courts have allowed civil claims against online platforms—despite Section 230's immunity—when a platform had a direct hand in creating the illegal user-generated content.

Thus, before Congress fundamentally changes Section 230, lawmakers should ask whether these bills are necessary to begin with.

Why Section 230 Matters

Section 230 is the part of the Telecommunications Act of 1996 that provides broad immunity to Internet intermediaries from liability for the content that their users create or post (i.e., user-generated content or third-party content).

Section 230 can be credited with creating today's Internet—with its abundance of unique platforms and services that enable a vast array of user-generated content. Section 230 has provided the legal buffer online entrepreneurs need to experiment with news ways for users to connect online—and this is just as important for today's popular platforms with billions of users as it is for startups.

Congress' rationale for crafting Section 230 is just as applicable today as when the law was passed in 1996: if Internet intermediaries are not largely shielded from liability for content their users create or post—particularly given their huge numbers of users—existing companies risk being prosecuted or sued out of existence, and potential new companies may not even enter the marketplace for fear of being prosecuted or sued out of existence (or because venture capitalists fear this).

This massive legal exposure would dramatically change the Internet as we know it: it would not only thwart innovation in online platforms and services, but free speech as well. As companies fall or fail to be launched in the first place, the ability of all Internet users to speak online would be disrupted. For those companies that remain, they may act in ways that undermine the open Internet. They may act as gatekeepers by preventing whole accounts from being created in the first place and pre-screening content before it is even posted. Or they may over-censor already posted content, pursuant to very strict terms of service in order to avoid the possibility of any user-generated content on their platforms and services that could get them into criminal or civil hot water. Again, this would be a disaster for online free speech. The current proposals to gut Section 230 raise the exact same problems that Congress dealt with in 1996.

By guarding online platforms from being held legally responsible for what thousands or millions or even billions of users might say online, Section 230 has protected online free speech and innovation for more than 20 years.

But Congress did not create blanket immunity. Section 230 reflects a purposeful balance that permits Internet intermediaries to be on the hook for their users' content in certain carefully considered circumstances, and the courts have expanded upon these rules.

Section 230 Does Not Bar Federal Prosecutors From Targeting Criminal Online Platforms

Section 230 has never provided immunity to Internet intermediaries for violations of *federal criminal law*—like the federal criminal sex trafficking statute (18 U.S.C. § 1591). In 2015, Congress passed the SAVE Act, which amended Section 1591 to expressly include "advertising" as a criminal action. Congress intended to go after websites that host ads knowing that such ads involve sex trafficking. If these companies violate federal criminal law, they can be criminally prosecuted in federal court alongside their users who are directly engaged in sex trafficking.

In a parallel context, a federal judge in the Silk Road case correctly ruled that Section 230 did not provide immunity against federal prosecution to the operator of a website that hosted other people's ads for illegal drugs.

By contrast, Section 230 does provide immunity to Internet intermediaries from liability for user-generated content under *state criminal law*. Congress deliberately chose not to expose these companies to criminal prosecutions in 50 different states for content their users create or post. Congress fashioned this balance so that federal prosecutors could bring to justice culpable companies while still ensuring that free speech and innovation could thrive online.

However, SESTA and its House counterpart would expose Internet intermediaries to liability under state criminal sex trafficking statutes. Although EFF understands the desire of state attorneys general to have more tools at their disposal to combat sex trafficking, such an amendment to Section 230 would upend the carefully crafted policy balance Congress embodied in Section 230.

More fundamentally, it cannot be said that Section 230's current approach to criminal law has failed. A Senate investigation earlier this year and a recent *Washington Post* article both uncovered information suggesting that Backpage.com not only knew that their users were posting sex trafficking ads to their website, but that the company also took affirmative steps to help those ads get posted. Additionally, it has been reported that a federal grand jury has been empaneled in Arizona to investigate Backpage.com. Congress should wait and see what comes of these developments before it exposes Internet intermediaries to additional criminal liability.

Civil Litigants Are Not Always Without a Remedy Against Internet Intermediaries

Section 230 provides immunity to Internet intermediaries from liability for user-generated content under civil law—whether federal or state civil law. Again, Congress made this deliberate policy choice to protect online free speech and innovation.

Congress recognized that exposing companies to civil liability would put the Internet at risk even more than criminal liability because: 1) the standard of proof in criminal cases is "beyond a reasonable doubt," whereas in civil cases it is merely "preponderance of the evidence," making the likelihood higher that a company will lose a civil case; and 2) criminal prosecutors as agents of the government tend to exercise more restraint in filing charges, whereas civil litigants often exercise less restraint in suing other private parties, making the likelihood higher that a company will be sued in the first place for third-party content.

However, Section 230's immunity against civil claims is not absolute. The courts have interpreted this civil immunity as creating a *presumption* of civil immunity that plaintiffs can rebut if they have evidence that an Internet intermediary did not simply host illegal user-generated content, but also had a *direct hand* in creating the illegal content. In a seminal 2008 decision, the U.S. Court of Appeals for the Ninth Circuit in *Fair Housing Council v. Roommates.com* held that a website that helped people find roommates violated fair housing laws by "inducing third parties to express illegal preferences." The website had required users to answer profile questions related to personal characteristics that may not be used to discriminate in housing (e.g., gender, sexual orientation, and the presence of children in the home). Thus, the court held that the website lost Section 230 civil immunity because it was "directly involved with developing and enforcing a system that subjects subscribers

to allegedly discriminatory housing practices." Although EFF is concerned with some of the implications of the *Roommates.com* decision and its potential to chill online free speech and innovation, it is the law.

Thus, even without new legislation, victims of sex trafficking may bring civil cases against websites or other Internet intermediaries under the federal civil cause of action (18 U.S.C. § 1595), and overcome Section 230 civil immunity if they can show that the websites had a direct hand in creating ads for illegal sex. As mentioned above, a Senate investigation and a *Washington Post* article both strongly indicate that Backpage.com would not enjoy Section 230 civil immunity today.

SESTA and its House counterpart would expose Internet intermediaries to liability under federal and state civil sex trafficking laws. Removing Section 230's rebuttable presumption of civil immunity would, as with the criminal amendments, disrupt the carefully crafted policy balance found in Section 230. Moreover, victims of sex trafficking can already bring civil suits against the pimps and "johns" who harmed them, as these cases against the direct perpetrators do not implicate Section 230.

Therefore, the bills' amendments to Section 230 are not necessary—because Section 230 is not broken. Rather, Section 230 reflects a delicate policy balance that allows the most egregious online platforms to bear responsibility along with their users for illegal content, while generally preserving immunity so that free speech and innovation can thrive online.

Congress Doesn't Understand How Section 230 Works

Aaron Mackey, Staff Attorney

https://www.eff.org/deeplinks/2017/09/stop-sesta-congress-doesnt-understand-how-section-230-works

As Congress considers undercutting a key law that protects online free speech and innovation, sponsors of the bills don't seem to understand how Section 230 (47 U.S.C. § 230) works.

EFF opposes the Senate's Stop Enabling Sex Traffickers Act (S. 1693) ("SESTA") and its House counterpart, the Allow States and Victims to Fight Online Sex Trafficking Act (H.R. 1865). These bills would roll back Section 230, one of the most important laws protecting online free speech and innovation.

Section 230 generally immunizes Internet intermediaries from legal liability for hosting user-generated content. Many websites or services that we rely on host third-party content in some way—social media sites, photo and video-sharing apps, newspaper comment sections, and even community mailing lists. This content is often offensive when, for example, users make defamatory statements about others. With Section 230, the actual speaker is at risk of liability—but not the website or blog. Without Section 230, these intermediaries would have an incentive to review every bit of content a user wanted to publish to make sure that the content would not be illegal or create a risk of legal liability—or to stop hosting user content altogether.

But according to one of SESTA's sponsors, Sen. Rob Portman (R-Ohio), EFF and other groups' concerns are overblown. In a recent floor speech, Sen. Portman said:

They have suggested that this bipartisan bill could impact mainstream websites and service providers—the good actors out there. That is false. Our bill does not amend, and thus preserves, the Communications Decency Act's Good Samaritan provision. This provision protects good actors who proactively block and screen for offensive material and thus shields them from any frivolous lawsuits.

Sen. Portman is simply wrong that the bill would not impact "good" platforms. He's also wrong about how Section 230's "Good Samaritan" provision would continue to protect online platforms if SESTA were to become law, particularly because that provision is irrelevant to the massive potential criminal and civil liability that the bill would create for online platforms, including the good actors.

Section 230 Has Two Immunity Provisions: One Related to User-Generated Content and One Called the "Good Samaritan" Immunity

We want to be very clear here, because even courts get confused occasionally. Section 230 contains two separate immunities for online platforms.

The first immunity (Section 230(c)(1)) protects online platforms from liability for hosting user-generated content that *others* claim is unlawful. If Alice has a blog on WordPress, and Bob accuses Clyde of having said something terrible in the blog's comments, Section 230(c)(1) ensures that neither Alice nor WordPress are liable for Bob's statements about Clyde.

The second immunity (Section 230(c)(2)) protects online platforms from legal challenges brought by their *own users* when the platforms filter, remove, or otherwise edit those users' content. In the context of the above example, Bob can't sue Alice if she unilaterally takes down Bob's comment about Clyde. This provision explicitly states that the immunity is premised on actions the platforms take in "good faith" to remove offensive content, even if that content may be protected by the First Amendment. This second provision is what Sen. Portman called the "Good Samaritan" provision. (Law Professor Eric Goldman has a good explainer about Section 230(c)(2).)

When EFF and others talk about the importance of Section 230, we're talking about the first immunity, Section 230(c)(1), which protects platforms in their role as hosts of user-generated content. As described above, Section 230(c)(1) generally prevents people who are legally wronged by user-generated content hosted on a platform (for example, defamed by a tweet) from suing the platform. Importantly, Section 230(c)(1) contains *no "good faith" or "Good Samaritan" requirement*.

Rather, Section 230(c)(1) provides platforms with immunity based solely on how they function: if providers offer services that enable their users to post content, they are generally shielded from liability that may result from that content. Full stop. Platforms' motives in creating or running their services are thus irrelevant to whether they receive Section 230(c)(1) immunity for user-generated content.

Section 230's "Good Samaritan" Immunity Wouldn't Protect Platforms From the Liability for User-Generated Content That SESTA Would Create

Sen. Portman's comments suggest that the current proposals to amend Section 230 would not impact the law's "Good Samaritan" provision found in Section 230(c)(2). That is debatable but beside the point, and it unnecessarily confuses the impact SESTA and its House counterpart would have on online free speech and innovation.

Sen. Portman's comments are beside the point because SESTA would blow a hole in Section 230(c)(1)'s immunity by exposing online platforms to increased liability for

user-generated content in two ways: 1) it would cease to protect platforms from prosecutions under state criminal law related to sex trafficking; and 2) it would cease to protect platforms from claims brought by private plaintiffs under both federal and state civil laws related to sex trafficking.

Section 230's "Good Samaritan" immunity simply doesn't apply to lawsuits claiming that user-generated content is illegal or harmful. Section 230(c)(2)'s "Good Samaritan" immunity is *irrelevant* for platforms seeking to defend themselves from the new claims based on user-generated content that SESTA would permit.

In those newly possible criminal and civil cases, platforms would be unable to invoke Section 230(c)(1) as a defense, and Section 230(c)(2) would not apply. Sen. Portman's comments thus betray a lack of understanding regarding how Section 230 protects online platforms.

Additionally, Sen. Portman implies that should SESTA become law, as long as platforms operate in "good faith," they will not be liable should content related to sex trafficking appear on their sites. This is a dangerous misstatement of SESTA's impact. Rather than recognizing that SESTA creates massive liability for all online platforms, Sen. Portman incorrectly implies that Section 230 currently separates good platforms from bad when it comes to which ones can be held liable for user-generated content.

Sen. Portman's comments thus do little to alleviate the damaging consequences SESTA will have on online platforms. Moreover, they appear designed to mask some of the bill's inherent flaws.

Amendments to Federal Criminal Sex Trafficking Law Sweep Too Broadly

Sophia Cope and Aaron Mackey, Staff Attorneys

https://www.eff.org/deeplinks/2017/09/stop-sesta-amendments-federal-criminal-sex-trafficking-law-sweep-too-broadly

EFF opposes the Senate's Stop Enabling Sex Traffickers Act (S. 1693) ("SESTA"), and its House counterpart the Allow States and Victims to Fight Online Sex Trafficking Act (H.R. 1865). Not only would both bills eviscerate the immunity from liability for user-generated content that Internet intermediaries have under Section 230, the bills would also amend the federal criminal sex trafficking statute to sweep in companies who may not even be aware of what their users are doing.

As we recently explained, Section 230 has always had an express exemption for federal criminal law, meaning that Internet intermediaries can be prosecuted in federal court. Thus, federal prosecutors have always been able use the federal criminal sex trafficking statute (18 U.S.C. § 1591) to go after online platforms without running into Section 230 immunity.

SESTA and its House counterpart would amend Section 1591 to expand federal criminal liability for Internet intermediaries—increasing the ways they may be on the hook for what are essentially the crimes of their users. These changes are not only unnecessary in light of current law, such an expansion of intermediary liability would undermine the online free speech and innovation that all Internet users have come to expect and enjoy.

Congress Already Gave Federal Prosecutors the Ability to Target Culpable Online Platforms

With the SAVE Act of 2015, Congress amended Section 1591 to make "advertising" sex trafficking a crime. Congress intended to target both the pimps who post sex trafficking ads and the online platforms who host such ads, in particular, classified ad websites like Backpage.com.

We have not yet seen a prosecution under the SAVE Act, but it has been reported that prosecutors have empaneled a federal grand jury in Arizona to investigate Backpage.com. Now Congress wants to further expand federal criminal liability under Section 1591 in dangerous ways—without proof that such an expansion is necessary.

Currently, given the 2015 amendments, Section 1591 can be read as prohibiting two main crimes:

- First, it is a crime for a person or entity to "advertise" sex trafficking or to benefit financially from "participation in a venture" that has engaged in advertising sex trafficking, *knowing* that an ad reflects a sex trafficking situation. Added by the SAVE Act, this crime was intended to apply only to the culpable hosts of online sex trafficking ads—those individuals or companies who, in fact, *know* that the ads are for sex trafficking.
- Second, it is a crime for a person or entity to engage in certain activities (other than advertising) related to sex trafficking or to benefit financially from "participation in a venture" that has engaged in certain activities related to sex trafficking. The statute lists the activities for which criminal liability attaches (specifically, if a person: recruits, entices, harbors, transports, provides, obtains, maintains, patronizes, or solicits). For this second set of crimes, the statute permits a lower standard for the defendant's state of mind: a person or entity who engages in these activities or benefits financially from "participation in a venture" that engages in these activities, knowing *or in reckless disregard of the fact* that sex trafficking is involved. Thus, individuals or companies need not, in fact, *know* that a "venture" involves sex trafficking, only that they *should have known*.

Congress assigned the higher "knowledge" standard to advertising sex trafficking in the SAVE Act in light of civil libertarians' concerns that attaching criminal liability to advertising implicates First Amendment rights.

SESTA Would Dangerously Expand Federal Criminal Liability to Encompass Innocent Online Platforms

The Senate bill would amend Section 1591 by further defining "participation in a venture" to include any activity that "assists, supports, or facilitates" sex trafficking.

Therefore, the Senate bill creates a third crime under Section 1591(a)(2):

- It is a crime for a person or entity to benefit financially from "participation in a venture" that has assisted, supported, or facilitated sex trafficking, knowing *or in reckless disregard of the fact* that sex trafficking is involved. (The House bill has similar amendments.)

There are two problems with this amendment to Section 1591(a)(2).

First, the words "assists, supports, or facilitates" are extremely vague and broad. Courts have interpreted "facilitate" in the criminal context simply to mean "to make easier or less difficult," as in using a phone to help "facilitate" a drug deal. A huge swath of innocuous intermediary products and services would fall within these newly prohibited activities, given that online platforms by their very nature make communicating and publishing "easier or less difficult."

Second, persons or entities would be criminally liable under the bill's vague and broad terms even if they do not actually *know* that sex trafficking is happening—much less *intend* to assist in sex trafficking. This would expose innocent individuals and companies to federal criminal liability should their products or services be misused by sex traffickers.

This reasonable reading of SESTA carries dangerous implications for all Internet intermediaries, not just classified ad websites like Backpage.com, as well as brick-and-mortar companies.

Any company in the chain of online content distribution—whether ISPs, web hosting companies, websites, search engines, email and text messaging providers, or social media platforms would be swept up by these amendments to Section 1591. All of these companies come into contact with user-generated content—whether ads, emails, text messages, or social media posts—some of which might involve sex trafficking. And all of these services can be said to "assist, support, or facilitate" sex trafficking. For example, should a messaging app be used by the perpetrators of sex trafficking to communicate with each other, a federal prosecutor could argue that such a service assisted, supported, or facilitated sex trafficking. Thus, all of these companies would be criminally liable under Section 1591 if a jury concludes—not that the companies actually *knew* their services were "facilitating" sex trafficking—but that they were "reckless" and thus *should have known* that was true in a particular case.

Additionally, the new federal criminal liability in Section 1591 created by SESTA would not be limited to online platforms, given that Section 1591 currently is not limited to online platforms but instead applies to "whoever" participates in a venture. Thus, on the face of the bill, any individual or company that "assists, supports, or facilitates" sex trafficking, *in reckless disregard of the fact* that sex trafficking is happening, is open to federal criminal liability. While perhaps not Congress' intent, this language could swallow up an endless list of companies who may not, in fact, be aware of what their customers are doing. For example, if a sex trafficker used a legitimate package delivery service or bank in the course of his illicit dealings, would those entities have "facilitated" sex trafficking?

In summary, just because Internet intermediaries cannot invoke Section 230 immunity when faced with liability under federal criminal law, it does not follow that the federal criminal sex trafficking statute should be further amended—beyond what the SAVE Act did—to sweep in what may be innocent Internet intermediaries and hold them responsible for the sex trafficking crimes of their users.

Section 1591—as amended two years ago—gives the U.S. Department of Justice more than enough leeway to prosecute culpable online platforms for their role in sex trafficking.

Whose Voices Will SESTA Silence?

Elliot Harmon, Activist, and Jeremy Gillula, Senior Staff Technologist

https://www.eff.org/deeplinks/2017/09/stop-sesta-whose-voices-will-sesta-silence

In all of the debate about the Stop Enabling Sex Traffickers Act (SESTA, S. 1693), there's one question that's received surprisingly little airplay: under SESTA, what would online platforms do in order to protect themselves from the increased liability for their users' speech?

With the threat of overwhelming criminal and civil liability hanging over their heads, Internet platforms would likely turn to automated filtering of users' speech in a big way. That's bad news because when platforms rely too heavily on automated filtering, it almost always results in some voices being silenced. And the most marginalized voices in society can be the first to disappear.

Section 230 Built Internet Communities

The modern Internet is a complex system of online intermediaries—web hosting providers, social media platforms, news websites that host comments—all of which we use to speak out and communicate with each other. Those platforms are all enabled by Section 230, a law that protects platforms from some types of liability for their users' speech. Without those protections, most online intermediaries would not exist in their current form; the risk of liability would simply be too high.

Section 230 still allows authorities to prosecute platforms that break federal criminal law, but it keeps platforms from being punished for their customers' actions in federal civil court or at the state level. This careful balance gives online platforms the freedom to set and enforce their own community standards while still allowing the government to hold platforms accountable for criminal behavior.

SESTA would throw off that balance by shifting additional liability to intermediaries. Many online communities would have little choice but to mitigate that risk by investing heavily in policing their members' speech.

Or perhaps hire computers to police their members' speech for them.

The Trouble with Bots

Massive cloud software company Oracle recently endorsed SESTA, but Oracle's letter of supportactually confirms one of the bill's biggest problems—SESTA would effectively require Internet businesses to place more trust than ever before in automated filtering technologies to police their users' activity.

While automated filtering technologies have certainly improved since Section 230 passed in 1996, Oracle implies that bots can now filter out sex traffickers' activity with near-perfect accuracy without causing any collateral damage.

That's simply not true. At best, automated filtering provides tools that can aid human moderators in finding content that may need further review. That review still requires human community managers. But many Internet companies (including most startups) would be unable to dedicate enough staff time to fully mitigate the risk of litigation under SESTA.

So what will websites do if they don't have enough human reviewers to match their growing user bases? It's likely that they'll tune their automated filters to err on the side of extreme caution—which means silencing legitimate voices. To see how that would happen, look at the recent controversy over Google's PerspectiveAPI, a tool designed to measure the "toxicity" in online discussions. PerspectiveAPI flags statements like "I am a gay woman" or "I am a black man" as toxic because it fails to differentiate between Internet users talking about themselves and making statements about marginalized groups. It even flagged "I am a Jew" as more toxic than "I don't like Jews."

See the problem? Now imagine a tool designed to filter out speech that advertises sex trafficking to comply with SESTA. From a technical perspective, creating such a tool that doesn't *also* flag a victim of trafficking telling her story or trying to find help would be extremely difficult. (For that matter, so would training it to differentiate trafficking from consensual sex work.) If Google, the largest artificial intelligence (AI) company on the planet, can't develop an algorithm that can reason about whether a simple statement is toxic, how likely is it that *any* company will be able to automatically and accurately detect sex trafficking advertisements?

Despite all the progress we've made in analytics and AI since 1996, machines still have an incredibly difficult time understanding subtlety and context when it comes to human speech. Filtering algorithms can't yet understand things like the motivation behind a post—a huge factor in detecting the difference between a post that actually advertises sex trafficking and a post that criticizes sex trafficking and provides support to victims.

This is a classic example of the "nerd harder" problem, where policymakers believe that technology can advance to fit their specifications as soon as they pass a law requiring it to do so. They fail to recognize the inherent limits of automated filtering: bots are useful in some cases as an aid to human moderators, but they'll never be appropriate as the unchecked gatekeeper to free expression. If we give them that position, then victims of sex trafficking may be the first people locked out.

At the same time, it's also extremely unlikely that filtering systems will actually be able to stop determined sex traffickers from posting. That's because it's not currently technologically possible to create an automated filtering system that can't

be fooled by a human. For example, say you have a filter that just looks for certain keywords or phrases. Sex traffickers will learn what words or phrases trigger the filter and avoid them by using other words in their place.

Building a more complicated filter—say, by using advanced machine learning or AI techniques—won't solve the problem either. That's because all complex machine learning systems are susceptible to what are known as "adversarial inputs"—examples of data that look normal to a human, but which completely fool AI-based classification systems. For example, an AI-based filtering system that recognizes sex trafficking posts might look at such a post and classify it correctly—unless the sex trafficker adds some random-looking-yet-carefully-chosen characters to the post (maybe even a block of carefully constructed incomprehensible text at the end), in which case the filtering system will classify the post as having nothing to do with sex trafficking.

If you've ever seen a spam email with a block of nonsense text at the bottom, then you've seen this tactic in action. Some spammers add blocks of text from books or articles to the bottom of their spam emails in order to fool spam filters into thinking the emails are legitimate. Research on solving this problem is ongoing, but slow. New developments in AI research will likely make filters a more effective aid to human review, but when freedom of expression is at stake, they'll never supplant human moderators.

In other words, not only would automated filters be ineffective at removing sex trafficking ads from the Internet, they would also almost certainly end up silencing the very victims lawmakers are trying to help.

Don't Put Machines in Charge of Free Speech

One irony of SESTA supporters' praise for automated filtering is that *Section 230 made algorithmic filtering possible*. In 1995, a New York court ruled that because the online service Prodigy engaged in some editing of its members' posts, it could be held liable as a "publisher" for the posts that it *didn't* filter. When Reps. Christopher Cox and Ron Wyden introduced the Internet Freedom and Family Empowerment Act (the bill that would evolve into Section 230), they did it partially to remove that legal disincentive for online platforms to enforce community standards. Without Section 230, platforms would never have invested in improving filtering technologies.

However, automated filters simply cannot be trusted as the final arbiters of online speech. At best, they're useful as an aid to human moderators, enforcing standards that are transparent to the user community. And the platforms using them must carefully balance enforcing standards with respecting users' right to express themselves. Laws must protect that balance by shielding platforms from liability for their customers' actions. Otherwise, marginalized voices can be the first ones pushed off the Internet.